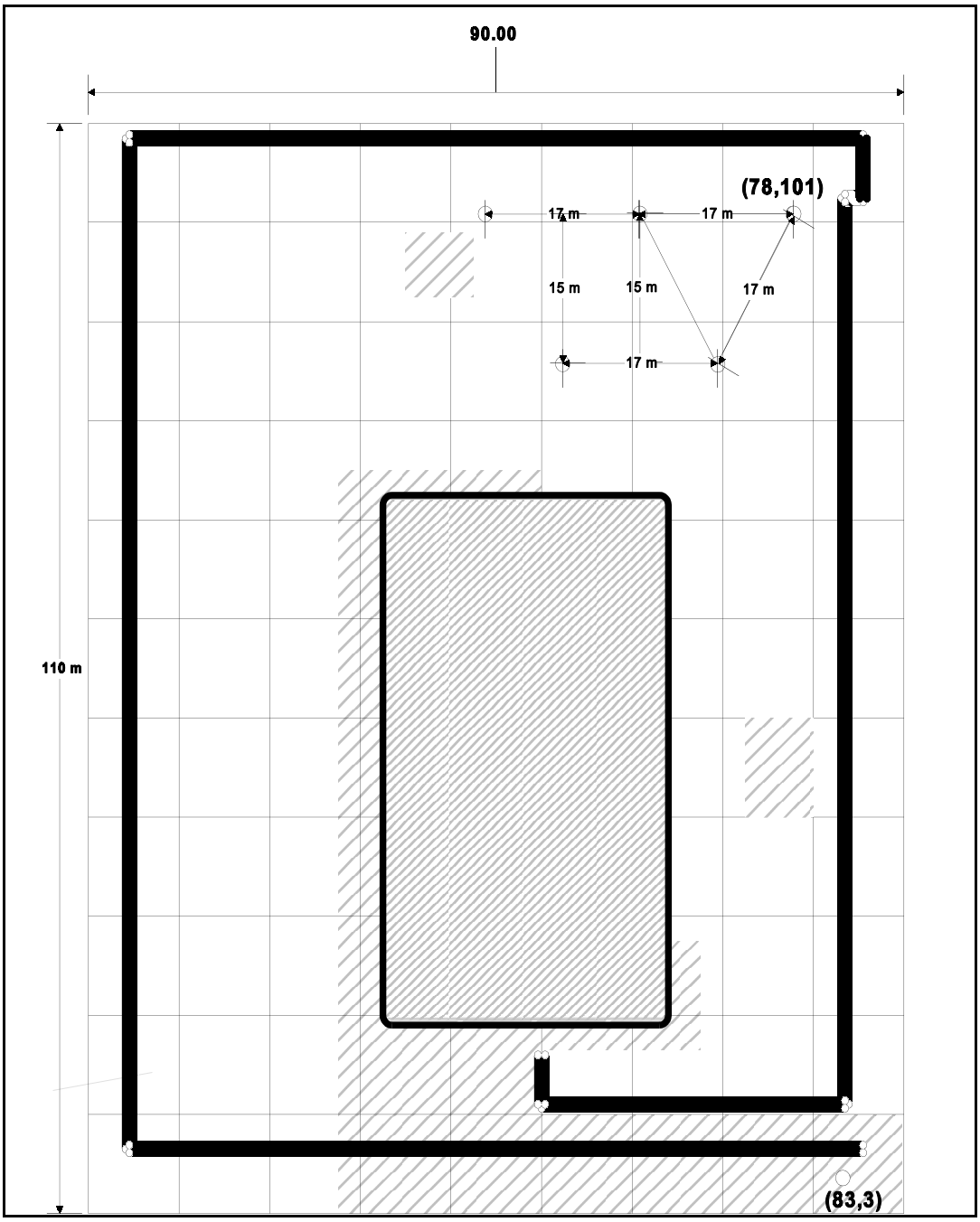

Figure 3.3 Laying Out a Triangular Grid

Using this procedure, the number of sampling points on the triangular grid within the sampling area may differ from the desired number, *n*, depending on the shape of the area. In this example, because of the irregular shape of the region caused by its wrapping around the building, 20 sampling points are found on the grid. If the number of points is greater than the desired number, use all the points.

If the number of points is less than the desired number, the additional points required may be identified using the same procedure that was used to determine the grid starting point. These will

be at individual random locations within the sampling area, and should be used regardless of where they occur relative to the grid. If more than a few random sample locations are needed, it is preferable to lower the grid spacing, *L*, and redraw the sampling grid.

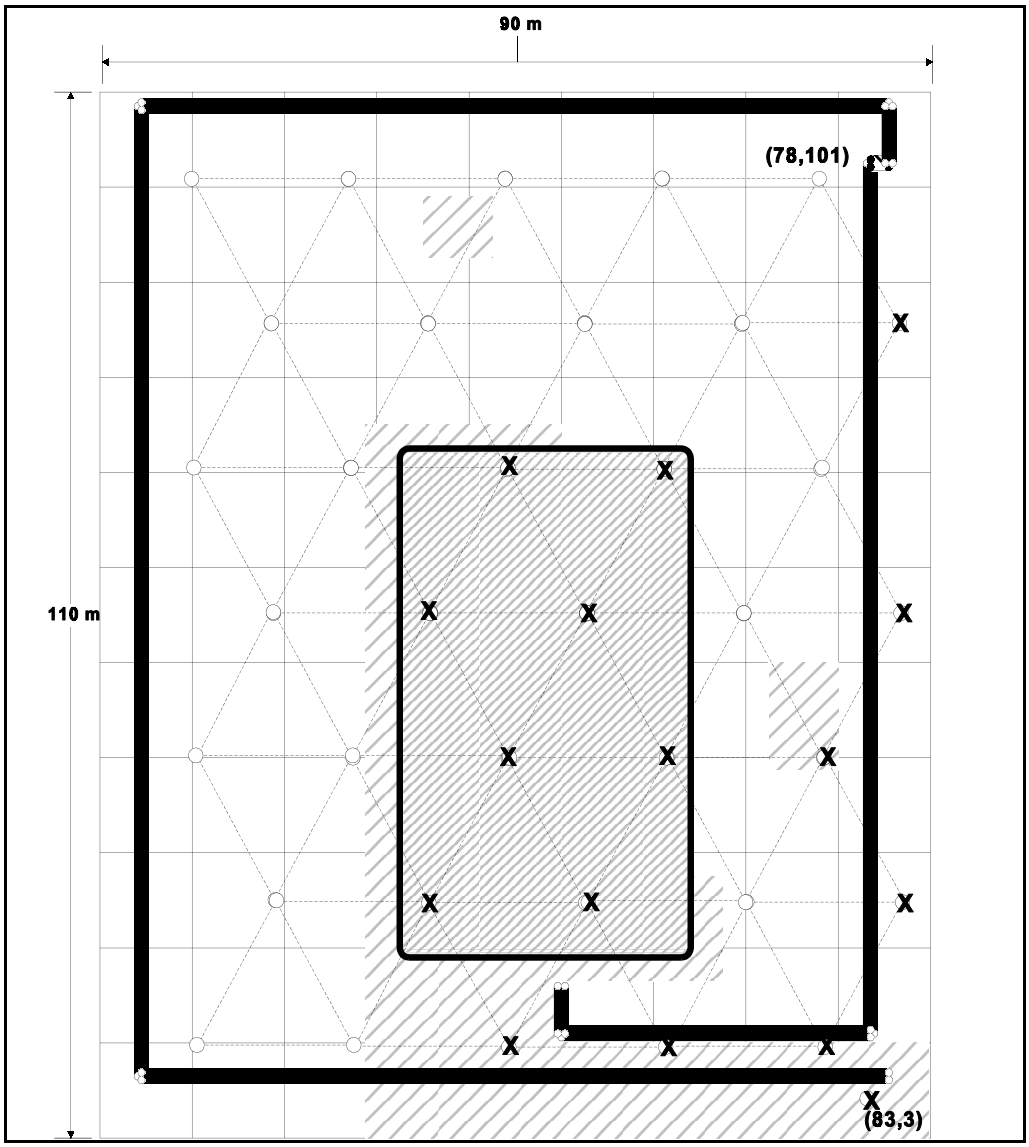

Figure 3.4 Completed Triangular Sampling Pattern

3.6 Develop a Decision Rule

A decision rule relates the concentration of residual radioactivity in the survey unit to the release criterion so that decisions can be made based on the results of the final status survey.

The decision rule proposed in this report for the final status survey consists of a statistical test and an elevated measurement comparison. The specific recommended statistical tests were discussed in Section 2.4, and the elevated measurement comparison was discussed in Section 2.5. Alternative statistical tests may be appropriate in specific circumstances, providing the assumptions of those tests are verified.

In some cases, it will not be necessary to formally conduct the statistical tests. If the radionuclide is not in background and radionuclide-specific measurements made, the survey unit meets the release criterion if all of the measurements are below the Derived Concentration Guideline Level for the mean residual radioactivity ($DCGL_W$) defined in Section 2.2.1. On the other hand, if the average of the measurements is above the $DCGL_W$, the survey unit does not meet the release criterion. It is only when some measurements are above the $DCGL_W$, but the average is below the $DCGL_W$, that the Sign test and the elevated measurement comparison need to be used.

If the radionuclide appears in background or if non-radionuclide-specific measurements made, the survey unit always meets the release criterion when the difference between the maximum survey unit measurement and the minimum reference area measurement is below $DCGL_{W.}$ If the difference between the survey unit and reference area averages is above $DCGL_W$, the survey unit fails to meet the release criterion. When the maximum difference is above the $DCGL_W$, but the average difference below $DCGL_W$, conduct the WRS test and elevated measurement comparison.

Recall that the $DCGL_W$ is the concentration level corresponding to the release criterion when the residual radioactivity is spread throughout the survey unit rather than is smaller elevated areas. The Lower Boundary of the Gray Region (LBGR) is the concentration level below which further remediation is not reasonably achievable. The null and alternative hypotheses for the statistical tests that were discussed in Section 2.3.1 can be restated in terms of the $DCGL_W$ and LBGR as follows.

Scenario A
Null Hypothesis:
H_0: The mean concentration of residual radioactivity above background in the survey unit exceeds the $DCGL_W$.
versus
Alternative Hypothesis:
H_a: The mean concentration of residual radioactivity above background in the survey unit does not exceed the LBGR.

Scenario B
Null Hypothesis:
H_0: The mean concentration of residual radioactivity in the survey unit is indistinguishable from background up to a level specified by the LBGR.
versus
Alternative Hypothesis:
H_a: The mean concentration of residual radioactivity in the survey unit distinguishable from background is in excess of the $DCGL_W$.

These hypotheses are stated in terms of the mean concentration, which is the parameter of interest. As discussed in Section 2.5, nonparametric tests are used to test these hypotheses, using the elevated measurement comparison to correct potential inaccuracies when the measurement distribution is very skewed.

The preceding paragraphs have indicated some decision rules for final status surveys. However,

there has been as yet no statement about how confident one can be that the decision is correct. While a formal statistical test may, in retrospect, not be needed to analyze the data, the survey must always be designed as if it were. Otherwise there is no basis for deciding the number of measurements to be taken, or with what precision. Nor will there be any basis for confidence in the resulting decision. This is the subject of the next Section. The hypothesis testing framework allows a estimate to be made of the Type I and Type II error probabilities. In addition, it is possible to similarly calculate the probability that the null hypothesis will be rejected (i.e., the power of the test) at any specific residual radioactivity concentration level, given some assumptions concerning the distribution of the residual radioactivity. In Section 3.7 on specifying the limits on decision errors, it will be seen that this allows the ALARA concept to be explicity incorporated into the decision-making process.

A different sort of decision rule is required for the elevated measurement comparison. As indicated in Section 2.2.1, the value of the $DCGL_{EMC}$ is based on a specific size area of elevated residual radioactivity. The area used during the survey planning to determine the $DCGL_{EMC}$ is based on the distance between the sampling locations on a systematic grid. The *actual* extent of an elevated area cannot be determined from a single measurement. When a measurement exceeds the $DCGL_{EMC}$, further investigation is required to determine both the size of the elevated area and its average concentration of residual radioactivity. Only then can it be determined if the TEDE due to the elevated area exceeds the release criterion. Thus, the decision rule for the EMC is a two stage process. In the first stage, areas are flagged as potentially elevated at specified investigation levels. In the second stage, the actual average concentration over the actual extent of elevated area is compared to the release criterion. The level at which measurements should be flagged depends on the survey unit classification, as discussed in Section 2.5.7.

3.7 Specify Limits on Decision Errors

A statistical decision error occurs when the null hypothesis is rejected when it is true (Type I) , or not rejected when it is false (Type II). The source of the uncertainty leading to such errors is the measurement variability, σ, discussed in Section 2.6. While the possibility of a decision error can never be totally eliminated, it can be controlled. Limits on decision errors are set to establish performance goals for the survey design. The two types of decision errors are classified as Type I and Type II decision errors, and were summarized in Table 2.2 for Scenario A and Scenario B. The probability of making a Type I decision error, or the level of significance, is called alpha (α). The probability of making a Type II decision error is called beta (β). The *power* of a test ($1-\beta$) is the probability of rejecting the null hypothesis when it is false.

This step in the DQO process is crucial. It is at this point that the limits on the decision errors rates are developed in order to establish appropriate goals for limiting uncertainty in the data. This is done by establishing the goals for the Type I error rate and the Type II error rate. The procedure for doing this follows.

3.7.1 Type I and Type II Decision Errors for Statistical Tests

A Type I error is made when the null hypothesis is rejected when it is true. A Type II error is made when the null hypothesis is not rejected when it is false. Thus the Type I and Type II errors have different meanings about meeting the release criteria, depending on whether Scenario A or

Scenario B is being considered. In an effort to avoid confusion, it will be convenient to say that a survey unit *passes* the statistical test if it is deemed to meet the release criterion as a result of that test. Otherwise, the survey unit will be said to *fail* the statistical test. The error rates can then be expressed as the probability that a survey unit passes when it should fail or fails when it should pass. This is summarized in Table 3.1. In Scenario A, the probability that a survey unit passes when it should fail is α and the probability it fails when it should pass is β. In Scenario B, the probability that a survey unit passes when it should fail is β and the probability it fails when it should pass is α. In Scenario A and B, the roles of α and β are reversed because the null and alternative hypotheses are reversed.

Table 3.1 Summary of Types of Decision Errors

If Survey Unit passes the statistical test...	**...when the True Condition of Survey Unit is...**	
	That It Does Not Meet Release Criterion	**That It Meets Release Criterion**
under Scenario A	Type I Error (Probability = α)	Correct Decision (Power = $1-\beta$)
under Scenario B	Type II Error (Probability = β)	Correct Decision (Probability = $1-\alpha$)

If Survey Unit fails the statistical test...	**...when the True Condition of Survey Unit is...**	
	That It Does Not Meet Release Criterion	**That It Meets Release Criterion**
under Scenario A	Correct Decision (Probability = $1-\alpha$)	Type II Error (Probability = β)
under Scenario B	Correct Decision (Power = $1-\beta$)	Type I Error (Probability = α)

Acceptable error rates can be established for either scenario by determining the desired probability for passing the survey unit as a function of the concentration of residual radioactivity actually remaining in the survey unit. Using the statistical tests recommended in this report, the probability that a survey unit passes decreases as the residual radiation concentration increases. At concentrations above background near the $DCGL_W$, this probability should be low in order to be adequately protective of public health. The probability that the survey unit passes should be high whenever the concentrations are near background in order to avoid unnecessary remediation costs. Somewhere in the range between residual radioactivity concentrations of zero and the $DCGL_W$ there is often is a concentration level such that remediation below this level is not considered to be reasonably achievable. considered unreasonable. The concentration range between this lower level and the $DCGL_W$, defines a *gray region* of residual radioactivity concentrations in which the consequences of decision errors are relatively minor. The

specification of a gray region is important because variability in the data may be such that a decision may be "too close to call" when the true but unknown value of the residual radioactivity concentration is very near the $DCGL_W$. The *Lower Boundary of the Gray Region,* the LBGR, is, by definition, the concentration value at which the acceptable probability of failing a survey unit when it should pass (β in Scenario A, α in Scenario B) is specified.

Specifying acceptable error rates is a means for bringing considerations of practicality directly into the decision making process. The probability limits assigned to points above and below the gray region should reflect the risks involved in making decision errors. These probabilities can then be converted to acceptable Type I and Type II error rates, α and β, using Table 3.1.

Figure 3.5 illustrates this process. For example, suppose it is considered that remediation to concentrations below one-half the $DCGL_W$ cannot be achieved with reasonable effort. The desired probability that the survey unit passes should then be set at a high value when the true residual radioactivity concentrations are at or below that level. This is the LBGR for this example. When the true concentration is at the $DCGL_W$, a small probability for passing the survey unit is desired. The line segments connecting the LBGR and $DCGL_W$ points with concentration values both higher and lower reflect the fact that the probability that the survey unit passes decreases with increasing concentration.

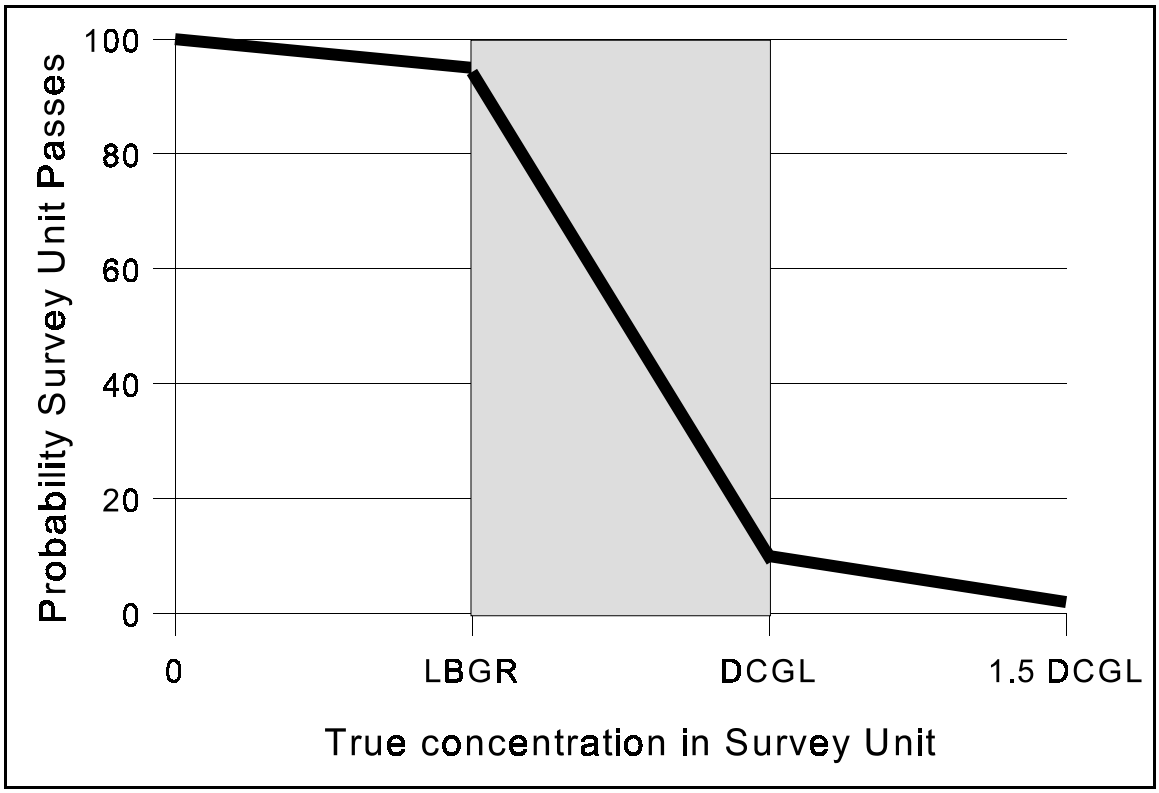

Figure 3.5 Example of Setting Acceptable Probabilities for Survey Unit Release

There is a relationship between α and β that is used in developing a survey design. When a fixed number of concentration measurements are made, increasing α generally decreases β, and decreasing α generally decreases β. Increasing the number of measurements allows either or both error rates to be decreased. Once the LBGR and $DCGL_W$ are specified, the number of

measurements needed to meet the desired values of α and β from the statistical test can be estimated using the estimated variance of the measured concentration distribution. This is discussed further in Section 3.8. The technical details are given in Chapter 9.

Constructing a curve such as that in Figure 3.5 is equivalent to specifying the desired power curve for the statistical test that is used to analyze the final status survey data. The desired power curve for the statistical test is selected during the DQO process by specifying the desired values for α and β at the lower (LBGR) and upper ($DCGL_W$) boundary of the gray region. By definition, the power of a test at any specific concentration is the probability that the null hypothesis is rejected when that is the true concentration in the survey unit. In conducting a statistical test, the value of the test statistic is calculated, and compared to a critical value. The critical value depends only on α and the number of measurements, *n*. The actual power may larger or smaller than the desired power at any specific value of the assumed true concentration. Thus, Figure 3.5 *is* the desired power for Scenario A, but it is one minus the power for Scenario B. More information about the power of the tests, how it is calculated, and a procedure for comparing the desired power to the actual power is given in Chapter 10.

The critical value of the test statistic depends, explicitly, only on α and the number of measurements, n. One consequence of this, is that the Type I error rate has traditionally been given precedence in experimental designs. Often, α is set at an arbitrarily low value, without regard to the impact on increasing β. EPA (QA/G-9, 1995) recommends that a more balance approach be used, that the errors rates, α and β, be considered simultaneously, and that several different sets of values be examined before finalizing the survey design. This is part of the optimization process discussed in Section 3.8.

Another consequence of the explicit dependence of the critical value on α, is the practice of calculating *p*-values, or levels of significance. Recall that α is the probability that the null hypothesis is rejected when it is actually true. No data are needed to calculate this probability. However, once the data are obtained, one may calculate the probability that a data set as extreme as that observed would occur when the null hypothesis is true. Unless this calculated "*p*-value" is greater than α, the null hypothesis is rejected. Difficulties can arise when a p-value very close to α is calculated. There is a tendency, in this case, to believe that one has "just missed" the desired result. There is also a temptation to believe that if more data are taken, the *p*-value will fall on the "correct" side of α. Unfortunately, unless the survey is designed specifically to be performed in two stages, the *p*-value calculated following the second stage of data collection will no longer be the correct one for the original null hypothesis. Some specific ways to construct two-stage tests are referenced in Chapter 14.

The value of α is fixed during the DQO process so that the critical value of the test statistic will be an objective standard of comparison for the measured data. This is necessary so that a clear line between pass and fail is drawn, despite the measurement uncertainty. In setting the value of α, it can be useful to consider what level of discomfort will be felt about the decision if the *p*-value that is observed should fall a little bit to either side of it. The time to adjust α and β is during the DQO process, not after the data are taken.

As stated earlier, the acceptable probabilities for survey unit release, and the corresponding values of α and β, that are set as goals during in the DQO process should reflect the risk involved

in making a decision error. The following are important considerations for this process:

- In radiation protection practice, the public health risk is modeled as a linear function of dose (BEIR, 1990). Therefore, a 10% change in dose results in a 10% change in risk. This situation is quite different from one in which there is a threshold. In the latter case, the risk associated with a decision error can be quite high, and a low decision error rate is desirable. When the risk is linear, higher error rates might be considered adequately protective at the boundaries of the gray region, especially since these errors are known to decrease as the concentration increases. One should consider, as part of the DQO process, the magnitude, significance, and potential consequences of decision errors at all concentration values. This is the purpose of the power curve.

- The DCGL itself has uncertainty associated with it. Since dose cannot be measured directly, dose pathway models are used. Many assumptions are made in converting doses to derived concentrations. To be adequately protective of public health, many models, especially screening models, are generally designed to guard against underestimating the dose that may be delivered by a given concentration of residual radioactivity. That is, the model assumptions tend to be such that the true dose delivered by residual radioactivity in the survey unit is very likely to be lower than that predicted by the model. Unfortunately, it is difficult to quantfy this. Nonetheless, it is probably safe to say that most models are conservative. This is an additional consideration that could support the use of higher acceptable error rates in some situations. The assumptions made in any model used to predict DCGLs for a site can be examined to determine if the use of site specific parameters result in large changes in the DCGLs, or whether a site-specific model should be developed rather than designing a survey around DCGLs that may be too conservative.

- The economic risk of requiring additional remediation when a survey unit already meets the release criterion may be highly non-linear. The costs will depend on whether the survey unit has already had remediation work performed on it, and the type of residual radioactivity present. There may be a threshold below which the remediation cost rises very rapidly. If so, a high probability of release is appropriate at that threshold value. This is primarily an issue for survey units that have a substantial likelihood of falling at or above the gray region for residual radioactivity. For survey units that are very lightly contaminated, or have been so thoroughly remediated that any residual radioactivity is expected to be far below the DCGL, smaller release probabilities may be tolerated, especially if final status survey sampling costs are a concern. Again, it is important to examine the probability that the survey unit passes over the entire range of possible residual radioactivity concentrations, below as well as above the gray region.

- Lower decision error rates may be possible if alternative sampling and analysis techniques can be used that result in higher precision. The same might be achieved with moderate increases in sample sizes. These alternatives should be explored before accepting higher design error rates. However, in some circumstances, such as high background variations, lack of a radionuclide specific technique, and/or radionuclides whose concentrations are very difficult and expensive to measure, error rates that are lower than the uncertainties in the dose estimates may be neither cost effective nor necessary for adequate radiation protection.